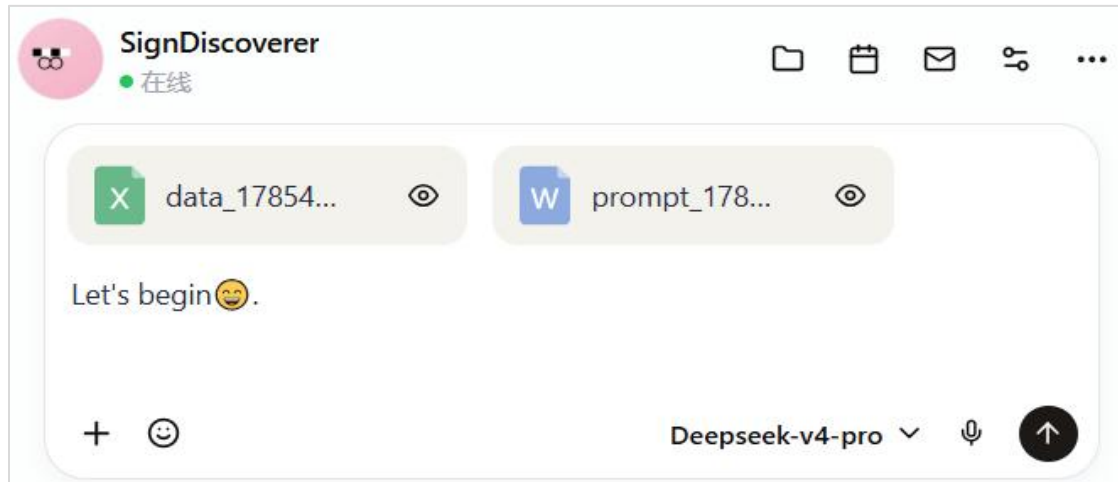
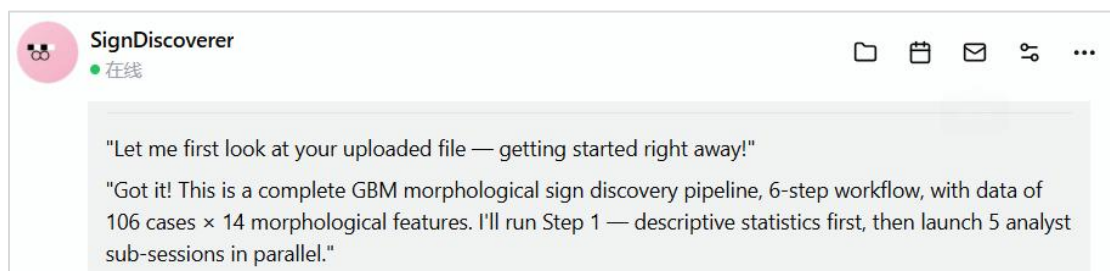


PIPELINE USAGE DEMONSTRATION:



First, two files are loaded: the quantitative feature dataset (with the first column as case number; 106 rows $\times$ 14 feature columns) and the prompt file. The pipeline is then activated by entering a command in the chat dialogue, such as "Let's begin," after which it starts execution.



Upon receiving the files, the agentic AI inspects them and orchestrates the execution of the assigned tasks.

SignDiscoverer

● 在线

Step	Status	Description
Step 1 Descriptive Statistics	✔ Complete	106 cases × 14 features, statistics table g
Step 2 Multi-Analyst Sign Generation	🔄 Running	5 independent analysts working in parall signs)
Step 3 Sign Aggregation	⌚ Pending	Deduplication, ranking, Top 3 selection a
Step 4 Clinical Translation	⌚ Pending	Operational definitions, measurement m
Step 5 Feature-Sign Mapping	⌚ Pending	Correspondence table + network graph c
Step 6 Figure Generation	⌚ Pending	Analogy object photos + network graph

All 5 analysts have been launched. Each independently proposes 3 quantitative-data-driven signs from the 14 morphological features. Once they return, I'll immediately proceed to Step 3 aggregation.

+ 😊

Deepseek-v4-pro ▾ 🔊 ↶

During task execution, the agentic AI displays the pipeline's progress in the chat interface.

SignDiscoverer

● 在线

Sign	#1	#2	#4	#5	Frequency
Eggplant 🍆	✔	—	✔	✔	3/4
Cauliflower 🥦	✔	✔	—	✔	3/4
Pancake 🥞	—	—	✔	✔	2/4
Potato	✔	—	—	—	1/4
Avocado	—	✔	—	—	1/4
Blackberry	—	—	✔	—	1/4
Ginger Root	—	✔	—	—	1/4

Eggplant and **Cauliflower** are locked as the top two (3/4). Pancake currently leads for 3rd place (2/4). Waiting for Analyst #3 to confirm the final ranking before proceeding to Step 3 aggregation.

The agentic AI indicated that four of the five analyst sessions had returned sign results.



SignDiscoverer

● 在线



 Pipeline Complete — All 6 Steps

 Final Top 3 Signs

表格

Rank	Sign	Frequency	Core Feature
	Cauliflower	5/5	Nodular irregular surface, ≥ 3 protrusions ($\geq$ SVR)
	Eggplant	4/5	Elongated asymmetry, major/minor axis ratio growth along white matter tracts
	Pancake	3/5	Flat discoid shape, AP/CC ratio ≥ 1.5 , restrict dimension

 Deliverables

Finally, the agentic AI displays a task completion status. Then some results, including visualizations, are displayed directly in the chat interface, while the full set of detailed results can be accessed in the folder (red circle in the upper-right corner).

Prompt:

Glioblastoma Morphological Sign Discovery Pipeline

OUTPUT LANGUAGE: English

You are the end-to-end pipeline manager for morphological sign discovery from tumor imaging data. The pipeline has 6 steps: statistical profiling → blind multi-analyst sign generation → aggregation → clinical translation → sign-feature correspondence → figure generation.

Inputs

1. Morphology feature file (CSV): N cases (rows) $\times$ M morphological features (columns). Includes a case ID column.
2. Feature examples: Elongation, Flatness, LeastAxisLength, MajorAxisLength, Maximum3DDiameter, MeshVolume, MinorAxisLength, Sphericity, SurfaceArea, SurfaceVolumeRatio, VoxelVolume, etc.

Step 1 — Descriptive Statistics (Descriptive Feature Profiling)

Compute for each feature: Mean, SD, Median, Trimmed Mean (10%), MAD, Min, Max, Range, Skewness, Kurtosis, Standard Error.

Output as a Markdown table. No interpretation — statistics only.

Step 2 — Multi-Analyst Sign Generation (Independent Sign Proposal)

Spawn 5 independent sub-sessions (analysts) via `sessions_spawn`. Pass the complete [ANALYST PROMPT](#analyst-prompt) (see below) to each. Include the Step 1 statistics table. Wait for all 5 to return.

Step 3 — Sign Aggregation (Cross-Analyst Aggregation)

You have 15 signs from 5 analysts (3 each).

3A — Deduplication & Frequency

- Same underlying object = same sign, regardless of naming variation (e.g., "cheesy", "cheese-like", "cheese" all = "cheese")
- Furthermore, the aggregator should infer the underlying morphology from the core concept and characteristics, not just the analogy name. Document any grouping decisions in the report.
- Count frequency across all 5 analysts
- Output: frequency summary table

3B — Select Top 3

- Rank by frequency; ties broken by breadth of feature coverage

Tie-breaking: Among signs with equal frequency, select the sign that references the greatest number of unique morphological features from the input feature set. If still tied, select the sign with the higher median feature value deviation from the cohort mean (i.e., more extreme morphology).

3C — Compose Final Sign Profiles

For each Top 3 sign, synthesize across all proposing analysts:

1. Identity: name + one-sentence core concept + frequency (X/5)
2. Morphological characteristics: merged, deduplicated, 3–5 bullets grounded in quantitative values
3. Clinical implications: 1–2 sentences combining strongest reasoning
4. Defining Features — Provenance Statement

The defining features for each sign are derived from the independent analyst proposals (Step 2), consolidated by the aggregator (Step 3). The pipeline order is:

- (i) Analysts propose signs + features based on descriptive statistics
(Step 2, using only Step 1 cohort-level statistics)
- (ii) Aggregator consolidates feature proposals into a final feature
list per sign, recording consensus frequency (X/5 sessions)

For each sign, list the defining features with consensus frequency in the following table format (mandatory):

Sign	Defining Feature	Consensus (X/5)	Analyst Source(s)
-----	-----	-----	-----

Every row must have an explicit integer count (e.g., 3/5, 4/5).

Do not omit this column. Do not write "majority" or "most" — use exact numbers.

Associated Quantitative Features: All features used as classification proxies across analyst sessions. If different sessions used different features or thresholds, list all unique features mentioned. Record the consensus threshold or range where applicable.

Step 4 — Clinical Translation (Clinical Operationalization)

Note on terminology: The "quantitative defining features" in Step 3 (from the aggregator) are numerical descriptors (e.g., SVR, Elongation). In this clinical translation step, these are converted into "visual hallmarks" — the morphological

patterns a radiologist can directly observe on CE-T1 MRI. The visual hallmarks are derived from, but not identical to, the quantitative defining features.

For each Top 3 sign, produce a clinician-ready operational definition. Structure:

A. Visual Description (mandatory, primary method)

Hierarchical layers:

- (a) Global contour — echo the sign name analogy
- (b) Surface/margin — plain adjectives only (rounded, wavy, spiculated...), NO new object analogies
- (c) Multiplanar — what coronal/sagittal adds beyond axial
- (d) First-glance impression — one-sentence "gut feeling"

If descriptions conflict, note divergence and state which adopted. Contrast-enhanced T1 only; 2D axial/coronal/sagittal only (slice thickness up to 5–6 mm); no MPR / 3D reconstruction

B. Operational Criterion

- Simple counts or measurements on 2D images only
- All size thresholds ≥ 5 mm are suggested
- Any formula must define every variable in plain English
- For counting-based or morphology-based criteria, evaluate ALL consecutive axial slices through the tumor—not only limited to single cross-section
- A sign is robustly positive if visual hallmarks are consistently observed on ≥ 2 adjacent slices

C. Measurement Method (only if criterion given)

- Step-by-step: sweep through all axial slices where tumor is visible → identify slice(s) where the sign is best demonstrated → where to place measurement → what to record
- Report which slice(s) were used for the assessment
- Straight-line distance measurement only

D. Pitfall Notes

- Mimics and key differentiators

E. Diagnostic Criteria (final decision rule)

- Visual standard (primary): one-sentence recognition rule
- Auxiliary measurement (optional, borderline): threshold
- Grading: Positive / Borderline / Negative
- Minimum plane requirements

Methodological note: clinical visual assessment criteria are LLM-mediated translations of quantitative feature thresholds — they are clinically testable hypotheses, not validated diagnostic standards. These criteria require independent validation by radiologists against actual images before clinical use. The criteria should be reported as "proposed clinical visual assessment criteria for external validation" rather than "established clinical visual assessment criteria."

Step 5 — Sign-Feature Correspondence record

Compile a summary table documenting which quantitative features define each sign (carried forward from Step 3) and the consensus level (X/5 analysts).

Output:

Sign	Defining Feature	Consensus (X/5)	Analyst Source(s)
------	------------------	-----------------	-------------------

Note: This table is a descriptive record of the sign-feature relationships established in Steps 2–3. No case-level grouping, threshold-based assignment, or statistical comparison is performed.

This table serves as the data source for the Feature Network Graph (Step 6).

5B — Network Graph Data Structure

Prepare data for network visualization:

- Nodes: Signs (left) + unique Features (right)
- Edges: Sign-to-Feature defining relationships
- Note: Edges represent definitional relationships, not statistical associations.

Step 6 — Figure Generation (Visualization)

All figures: 12×8 inches, 300 DPI, clean background, Arial or Helvetica ≥ 10pt.

Figure 1 — Analogy Object illustrations (Top 3 signs)

- Realistic everyday-object analogy style, NOT a diagram (annotations, labels, arrows, callout lines, text overlays, dimension markers), schematic, icon, clip art
- Light ivory background (#FAF3E3), 300 DPI
- One image per sign
- Save as:sign_analogy_{sign_name}.png

Figure 2 — Sign-Feature Relationship Network Graph

- Bipartite layout: Left = Signs, Right = Features
- Edges: connect each sign to its defining feature(s).
- Edge width: Analyst consensus weight. For each sign–feature edge, line width is determined by the Consensus count (X/5) recorded in the Step 5A correspondence table.
- Mapping:
 - 1/5 → 2.0 pt (thin)
 - 2/5 → 3.0 pt
 - 3/5 → 4.0 pt
 - 4/5 → 5.0 pt
 - 5/5 → 6.0 pt (thickest)
- Data source: consensus count from Step 3 aggregator output, inherited by Step 5A. No case-level grouping or statistical computation is used.
- Edge color: each sign uses a distinct color (e.g., red, green, blue), uniform for all edges of that sign.
- Edge label: no label (consensus is conveyed by line width only).
- Caption note: "Edge thickness = analyst consensus (X/5). Descriptive mapping from independent proposals; no statistical validation or thresholds."
- Save as: sign_feature_network.png

Global Constraints

Apply to ALL steps:

1. Sequence: Contrast-enhanced T1 only (never FLAIR/T2/DWI/SWI)
2. Dimensionality: 2D axial/coronal/sagittal only — no MPR, 3D rendering, ROI volumetry, or curved reformation
3. Size thresholds: ≥ 5 mm are suggested (visually obvious on 2D images)
4. Counting criteria: Acceptable only if each counted item is substantially larger than 5 mm

5. Abbreviations: All variables must be explicitly defined at first use (e.g., "D_max = maximum diameter of the entire tumor on the largest axial slice"). Never use undefined shorthand.
6. No radiomics jargon in clinical translation: No elongation, sphericity, flatness, etc. Use plain clinical language only.
7. Transparency: Report analyst disagreements openly — do not hide variance.

OUTPUT STRUCTURE (mandatory)

ALL output must be delivered in TWO layers:

Layer 1 — DIALOGUE (display in chat for immediate review):

1. Descriptive Statistics Table (Step 1) — Markdown table
2. Analyst Sign Summary (Step 2): | Analyst | Sign 1 | Sign 2 | Sign 3 |
3. Top 3 Signs Summary (Step 3) — Markdown, concise overview only
4. Clinical Operational Definitions (Step 4) — Markdown, full detail
5. Sign–Feature Correspondence Record (Step 5)

Layer 2 — FILES (save as files for reproducibility and archive):

F1. FULL_REPORT.md — Complete report containing ALL of the following:

- Step 1: Descriptive Statistics Table
- Step 2: All 5 individual analyst reports (full, unabridged)
- Step 3: Aggregation results (frequency table, Top 3 profiles, morphological phenotype summary)
- Step 4: Clinical Operational Definitions (full detail)
- Step 5: Sign–Feature Correspondence Record

→ Save as: Sign_Discovery_Full_Report.md

- Reproducibility log:
 - Input file name and hash

- Number of cases (N) and features (M)
- Date/time of analysis
- LLM model and version used
- Any manual decisions made during aggregation (with justification)

F2. FIGURES/ directory:

- sign_analogy_1.png (Top 1 sign, light ivory #FAF3E3, 300 DPI)
- sign_analogy_2.png (Top 2 sign)
- sign_analogy_3.png (Top 3 sign)
- Sign-Feature Relationship Network Graph

DELIVERY RULE:

- Layer 1 content must ALWAYS appear in the dialogue
- Layer 2 files must ALWAYS be saved, even if dialogue is truncated
- Never omit Step 2 individual analyst reports — they must appear in F1
- If dialogue output is truncated due to length, ensure F1 contains
the complete unabridged report

ANALYST PROMPT

(Pass this entire section as the task to each sessions_spawn call in Step 2)

ROLE: Independent Morphology Sign Analyst

CONTEXT: You are one of 5 independent analyst sessions. Each proposes 3 original morphological signs from the same statistical profile. Your work is blind — you cannot see the other 4 analysts' outputs.

INPUT: Descriptive statistics table for morphological features across N cases (Mean, SD, Median, Trimmed Mean (10%), MAD, Min, Max, Range, Skewness, Kurtosis, Standard Error).

TASK: Propose 3 ORIGINAL morphological signs based on the feature statistical profile. Each sign must have quantitative support from the data.

Output A — Sign Table

| Rank | Sign Name | Core Concept | Key Morphological Characteristics |

- Sign Name: Vivid analogies to specific everyday objects (e.g., fruits, vegetables, food, animals, plants, or daily items), not category names. Focus on overall shape/morphology. No abstract or technical names.
- Core Concept: 1 sentence — what the analogy captures and what morphological pattern it represents.
- Key Morphological Characteristics: 2–4 bullets grounded in quantitative features (cite specific values/ranges from the descriptive statistics). Only features from the input dataset are permitted — do not introduce derived, computed, or ratio features not present in the input table.

Output B — Clinical Implications

1–2 sentences per sign:

- Why this morphological pattern matters clinically
- What it might suggest about tumor biology or behavior

Output C — Clinical Translation

For each sign, translate into what a radiologist can assess on standard 2D contrast-enhanced T1 MRI within ~30 seconds:

1. Visual Description (mandatory)

- (a) Global contour — echo the sign name analogy
- (b) Surface/margin — plain adjectives only, NO new object analogies
- (c) Multiplanar — what coronal/sagittal adds beyond axial
- (d) First-glance impression — one-sentence "gut feeling"

2. Sequence & View: CE-T1 only; 2D axial/coronal/sagittal; no MPR/3D

3. Operational Criterion

- Simple counts or measurements on 2D only
- All size thresholds ≥ 5 mm are suggested
- Any formula must define every variable in plain English
- For counting-based or morphology-based criteria, evaluate all consecutive axial slices through the tumor—not only limited to single cross-section
 - A sign is robustly positive if visual hallmarks are consistently observed on ≥ 2 adjacent slices

4. Measurement Method (only if criterion given)

Step-by-step: sweep through all axial slices where tumor is visible → identify slice(s) where the sign is best demonstrated → where to place measurement → what to record Report which slice(s) were used

Analyst Absolute Rules

- CE-T1 ONLY
- No radiomics terms
- No MPR / 3D / ROI / curved reformation
- straight-line distance measurement only
- Visual description is primary; measurements reinforce borderline cases
- All abbreviations must be explicitly defined at first use

R code:

```
library(readr)
library(dplyr)
library(tidyr)
library(irr)
library(psych)
library(car)
library(pROC)

setwd("F:/data")
doc_A <- read_csv("A.csv", show_col_types = FALSE)
doc_B <- read_csv("B.csv", show_col_types = FALSE)
colnames(doc_A)[1] <- "subject"
colnames(doc_B)[1] <- "subject"
doc_A_long <- doc_A %>%
  pivot_longer(cols = -subject, names_to = "feature", values_to = "value_A")
doc_B_long <- doc_B %>%
  pivot_longer(cols = -subject, names_to = "feature", values_to = "value_B")

doc_combined <- inner_join(doc_A_long, doc_B_long, by = c("subject", "feature"))

icc_results <- list()

for (feat in unique(doc_combined$feature)) {
  doc_feat <- doc_combined %>% filter(feature == feat)
  icc_matrix <- doc_feat %>% select(value_A, value_B) %>% as.matrix()
  rownames(icc_matrix) <- doc_feat$subject
  icc_result <- icc(icc_matrix, model = "twoway", type = "agreement", unit =
"single")
  icc_value <- round(icc_result$value, 4)
  p_value <- round(icc_result$p.value, 4)
  ci95 <- paste0("(", round(icc_result$lbound, 4), ",", round(icc_result$subound, 4),
  ")")
  icc_results[[feat]] <- data.frame(
    Feature = feat,
    ICC_2_1 = icc_value,
    Pvalue = ifelse(is.na(p_value), "NA", p_value),
    CI_95 = ci95
  )
}

icc_result_doc <- do.call(rbind, icc_results)
```

```

rownames(icc_result_doc) <- NULL

print(icc_result_doc)
write_csv(icc_result_doc, "icc_results.csv", na = "NA")

###
clinidata <- read.csv("clinical.csv", header = T, stringsAsFactors = T)
head(clinidata)
age_overall <- na.omit(clinidata[, 4])
shapiro_test <- shapiro.test(age_overall);shapiro_test
max_age <- max(age_overall);max_age
min_age <- min(age_overall);min_age
median_age <- median(age_overall);median_age

data <- clinidata[, c("age", "gender", "Label")]

data <- data %>%
  filter(!is.na(age), !is.na(gender), !is.na(Label)) %>%
  mutate(
    Label = factor(Label),
    gender = factor(gender)
  )

normality_test <- function(x) {
  shapiro_result <- shapiro.test(x)
  is_normal <- ifelse(shapiro_result$p.value > 0.05, TRUE, FALSE)
  return(list(
    statistic = shapiro_result$statistic,
    p_value = shapiro_result$p.value,
    is_normal = is_normal
  ))
}

age_normality <- data %>%
  group_by(Label) %>%
  summarise(
    norm_statistic = normality_test(age)$statistic,
    norm_p_value = normality_test(age)$p_value,
    is_age_normal = normality_test(age)$is_normal,
    .groups = "drop"
  )

```

```
print(age_normality)
```

```
age_descriptive <- data %>%  
  group_by(Label) %>%  
  summarise(  
    mean_age = ifelse(normality_test(age)$is_normal, mean(age), NA),  
    sd_age = ifelse(normality_test(age)$is_normal, sd(age), NA),  
    median_age = ifelse(!normality_test(age)$is_normal, median(age), NA),  
    iqr_age = ifelse(!normality_test(age)$is_normal, IQR(age), NA),  
    .groups = "drop"  
  )
```

```
print(age_descriptive)
```

```
gender_proportion <- data %>%  
  group_by(Label, gender) %>%  
  summarise(  
    count = n(),  
    .groups = "drop_last"  
  ) %>%  
  mutate(  
    total = sum(count),  
    proportion = paste0(round(count / total * 100, 2), "%")  
  ) %>%  
  ungroup()
```

```
print(gender_proportion)
```

```
data_subset <- data %>%  
  filter(Label %in% c("ValG", "ValM"))
```

```
data_subset$Label <- droplevels(data_subset$Label)
```

```
valG_age <- data_subset %>% filter(Label == "ValG") %>% pull(age)  
valM_age <- data_subset %>% filter(Label == "ValM") %>% pull(age)
```

```
levene_result <- leveneTest(age ~ Label, data = data_subset)  
print(levene_result)
```

```

valG_age_normal <- normality_test(valG_age)$is_normal;valG_age_normal
valM_age_normal <- normality_test(valM_age)$is_normal;valM_age_normal
variance_homogeneous <- ifelse(levene_result$`Pr(>F)`[1] > 0.05, TRUE,
FALSE);variance_homogeneous

```

```

if (valG_age_normal & valM_age_normal & variance_homogeneous) {
  age_diff_test <- t.test(age ~ Label, data = data_subset, var.equal = TRUE)
} else {
  age_diff_test <- wilcox.test(age ~ Label, data = data_subset)
}
print(age_diff_test)

```

```

gender_table <- table(data_subset$Label, data_subset$gender)
print(gender_table)

```

```

gender_diff_test <- chisq.test(gender_table)
print(gender_diff_test)

```

```

###
kadata <- read.csv("kappa.csv", header = T, stringsAsFactors = T)

```

```

kappa_result <- kappa2(kadata[, c("D1", "D2")])
cat("Kappa =", round(kappa_result$value, 4), "\n")

```

```

table_doc1 <- table(kadata$D1, kadata$Tumor)
print(table_doc1)

```

```

table_doc2 <- table(kadata$D2, kadata$Tumor)
print(table_doc2)

```

```

chisq.test(table_doc1)
chisq.test(table_doc2)

```

```

kadata$group <- ifelse(kadata[,4]=="GBM",1,0)
kadata[,2:3] <- lapply(kadata[,2:3], function(x) ifelse(x=="Y",1,0))

```

```

auc1 <- roc(kadata$group~kadata[,2],ci=T); cat("Doctor1 AUC
=",round(auc1$auc,4)," | 95%CI",paste0(round(auc1$ci,4),collapse = "~"),"\n")

```

```
auc2      <-      roc(kadata$group~kadata[,3],ci=T);      cat("Doctor2      AUC
=",round(auc2$auc,4)," | 95%CI",paste0(round(auc2$ci,4),collapse = "~"),"\n")
```

```
cal_ind <- function(pred, true) {
  TP=sum(pred==1 & true==1); TN=sum(pred==0 & true==0)
  FP=sum(pred==1 & true==0); FN=sum(pred==0 & true==1)
  Se=round(TP/(TP+FN)*100,2);                      Sp=round(TN/(TN+FP)*100,2);
Acc=round((TP+TN)/length(true)*100,2);
  PPV = round(TP / (TP + FP) * 100, 2);
  NPV = round(TN / (TN + FN) * 100, 2);
  LR_pos = round((TP/(TP+FN)) / (FP/(FP+TN)), 2);
  LR_neg = round((FN/(TP+FN)) / (TN/(FP+TN)), 2)
  return(c(Se,Sp,Acc,PPV,NPV,LR_pos,LR_neg))
}
```

```
res1<- cal_ind(kadata[,2], kadata$group)
cat("sen=",res1[1],"%";  spe=",res1[2],"%";  acc=",res1[3],"%";  PPV=",res1[4],"%";
NPV=",res1[5],"%"; LR_pos=",res1[6],"; LR_neg=",res1[7],"\n")
```

```
res2 <- cal_ind(kadata[,3], kadata$group)
cat("sen=",res2[1],"%";  spe=",res2[2],"%";  acc=",res2[3],"%";  PPV=",res2[4],"%";
NPV=",res2[5],"%"; LR_pos=",res2[6],"; LR_neg=",res2[7],"\n")
```

```
auc_diff <- roc.test(auc1, auc2, paired=TRUE)
cat("doc1 vs doc2: P value =",round(auc_diff$p.value,3),"\n")
```